

v1 总结



项目小结（v1 版本）



当前完成度

本次 v1 版本在 v0 基础上进行了较为完整的拓展和调整，主要完成了以下工作：

- 对 v0 中的代码与逻辑进行了重构和精简，统一格式与风格；
- 针对模型部分进行了更全面的探索与测试，涵盖 Logistic Regression、Random Forest、MLP、HistGBR 等多个模型；
- 在 O1/O2 中尝试引入 MLP 和 Gradient Boosting 等新模型；
- 明确对部分模型如 KNN 和 Stepwise 进行尝试后放弃，原因包括：
 - KNN 效果不明显且训练时间极长；
 - Stepwise 逻辑较复杂，运行缓慢且难以控制变量选择过程，最终被排除；
- 所有任务均使用了独立测试集（proj-test-data.csv）验证最终模型，保证泛化能力评估的完整性；
- 报告支持中实现中葡英三语支持，并整理了不同版本以供小组选择提交形式。（如果还没有的话见谅，可以用沉浸式翻译插件进行pdf翻译）



已知问题与待优化部分

- 尽管最终内容基本完善，但优化过程拖延了很长时间，期间没有持续保持专注，任务推进节奏断断续续；
- 一些 notebook 结构未能保持良好一致性（例如命名风格、注释规范、重复处理逻辑等），整体结构仍有提升空间；
- 个别数据处理部分（如 EDA 中的深层分析）受时间限制未能展开得更深入；
- 对模型解释性提升（如 SHAP/LIME）没有覆盖，未来可以作为补充拓展。




小结与感想


这一版本在思路上比 v0 更加完整，尝试面也更广，覆盖了多个主流模型与调参策略，较好地体现了理论课中的知识点。

但反过来看，也暴露出自身在项目管理和自律方面的不足——中期缺乏节奏控制，导致许多优化任务被压到最后完成；代码虽然可用，但缺乏模块性和一致性，也提醒自

己在未来的项目中需要更多注重结构与复用。

整体而言，这是一次务实且收获丰富的实战训练。接下来，希望自己能在项目组织与代码维护上进一步提升，在未来更大规模的项目中表现得更稳健。

 当前版本：v1

 报告撰写日期：2025年5月23日